Adaptive Design for Clinical Trials

Mark Chang

Millennium Pharmaceuticals, Inc., Cambridge, MA 02139,USA
(e-mail: Mark.Chang@Statisticians.org)

Abstract. Adaptive design is a trial design that allows modifications to some aspects of the trial after its initiation without undermining the validity and integrity of the trial. Adaptive design makes it possible to discover and rectify inappropriate assumptions in trial designs, lower development costs and reduce the time to market. It has become very attractive to the pharmaceutical industries. In this paper, adaptive designs for clinical trials with multiple endpoints including binary, ordinal, normal, and survival responses are studied using computer simulations.
Keywords: Adaptive design, Sequential design, Adaptive randomization.

1 Overview of Adaptive Design

Drug development is a sequence of complicated decision-making processes. Options are provided at each stage and decisions are dependent on the prior information and the probabilistic consequence of each action (decision) taken. This requires the trial design to be flexible such that it can be modified during the trial process. Adaptive design emerges for this reason and has become very attractive to pharmaceutical industries. An adaptive design is a design that allows modifications to some aspects of the trial after its initiation without undermining the validity and integrity of the trial. The following are the examples of modifications to a trial.

- Sample size re-estimation
- Early stopping due to efficacy or futility
- Adaptive randomization
- Dropping inferior treatment groups

There are several methods available for adaptive designs such as the Fisher's combination of independent p-values [Bauer and Kohne, 1994], Brownian motion [Lan and Demets, 1988] [Lin et al., 1999], conditional power approach [Proschan and Hunsberger 1995], [Babb and Rogatko, 2004], and approach using down-weighting later-stage data [Cui et al., 1999] have been used in group sequential and adaptive designs. However, in this paper, we will discuss the use of computer trial simulation (CTS) for adaptive design. CTS provides a unique and powerful tool for achieving the optimal design. An overall process of an adaptive design is depicted in figure 1.

[0] Invited paper for International Symposium on Applied Stochastic Models and Data Analysis, France, May 2005.

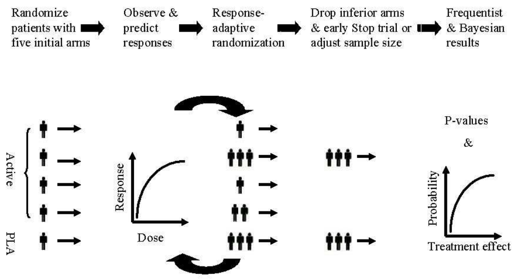

Randomize patients with five initial arms
Observe & predict responses
Response-adaptive randomization
Drop inferior arms & early Stop trial or adjust sample size
Frequentist & Bayesian results
Active
PLA
Response
Dose
P-values
&
Probability
Treatment effect

Fig. 1. Overview of Adaptive Design.

2 Utility-Based Trial Objective

A clinical trial typically involves multiple endpoints such as efficacy, safety and cost. Therefore a single measure, i.e., utility index, which summaries the effects of major endpoints is desirable. The trial objective then becomes to find the dose or treatment with the maximum response probability (rate). The response probability is defined as $Pr(u >= c)$ where u is utility index and c is a threshold. The utility index is the weighted average of trial endpoints such as safety and efficacy. The weights and the threshold are often determined by experts in the relevant field.

3 Dose-Response Model

The response of an ongoing trial can be modeled using a function. We find that the following so-called hyper-logistic function can be used model many different response shapes. The hyper-logistic function is defined by the probability of response

$$\Pr(x = 1) = (a_1 \exp(a_2 x) + a_3 \exp(-a_4 x))^{-a_5}$$

The modeling can be on a continual basis, i.e., the model is updated when new response data become available. This approach refers to the continual re-assessment method (CRM), which can be either Bayesian or frequentist based method. If the observed the responses are used as the basis for an adaptation instead of modeled or predicted response, we call it null-model approach.

4 Adaptation Rules

4.1 Randomization Rules

It is desirable to randomize more patients to superior treatment groups. This can be accomplished by increasing the probability of assigning a patient to the treatment group when the evidence of responsive rate increases in a group. The response-adaptive randomization rule can be Randomized-Play-the-Winner (RPW) [Rosenberger and Lachin, 2002], or Utility offset model.

Response-adaptive randomization requires unblinding the data, which may not feasible at real time. There is often a delayed response, i.e., randomizing the next patient before knowing responses of previous patients. Therefore, it is practical to unblind the data several times during the trial, i.e., group sequential response-adaptive randomization, instead of fully sequential adaptive randomization.

4.2 Early Stopping Rules

It is desirable to stop trial when the efficacy or futility of the test drug becomes obvious during the trial. To stop a trial prematurely, we provide a threshold for the number of subjects randomized and at least one of the following:

(1) Utility rules: The difference in response rate between the most responsive group and the control group exceeds a threshold and the corresponding two-sided 95% naïve confidence interval lower bound exceeds a threshold.

(2) Futility rules: The difference in response rate between the most responsive group and the control is lower than a threshold and the corresponding two-sided 90% naïve confidence interval upper bound is lower a threshold.

4.3 Rules for Dropping Losers

In addition to the response-adaptive randomization, you can also improve the efficiency of a trial design by dropping some inferior groups (losers) during the trial. To drop a loser, we provide two thresholds for (1) maximum difference in response rate between any two dose levels, and (2) the corresponding two-sided 90% naïve confidence lower bound. We may choose to retain all the treatment groups without dropping a loser, and/or to retain the control group with a certain randomization rate for the purpose of statistical comparisons between the active groups and the control.

4.4 Sample Size Adjustment

Sample size determination requires anticipation of the expected treatment effect size defined as the expected treatment difference divided by its standard deviation. It is not uncommon that the initial estimation of the effect size turns out to be too large or small, which consequently leads to an underpowered or overpowered trial. Therefore, it is desirable to adjust the sample size according to the effect size for an ongoing trial.

The sample size adjustment is determined by a power function of treatment effect size, i.e.,

$$N = N_0 \left(\frac{E_{0\,\max}}{E_{\max}} \right)^a \tag{1}$$

where N is the newly estimated sample size, N_0 the initial sample size, and a a constant. The effect size $E_{\max}$ is defined as

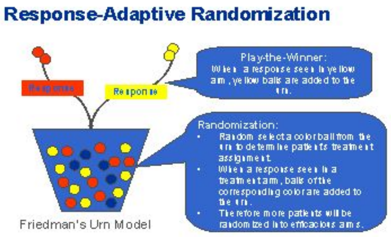
$$E_{\max} = \frac{p_{\max} - p_1}{\sigma^2}; \;\; \sigma^2 = \bar{p}(1 - \bar{p}); \;\; \bar{p} = \frac{p_{\max} + p_1}{2};$$

$p_{\max}$ and p_1 are the maximum response rates, respectively, and the control response rate, and $E_{0\,\max}$ is the initial estimation of $E_{\max}$.

5 Response-Adaptive Randomizations

The conventional randomization refers to any randomization procedure with a constant treatment allocation probability such as simple randomization. Unlike the conventional randomization, response-adaptive randomization is a randomization in which the probability of allocating a patient to a treatment group is based on the response of the previous patients. The purpose is to improve the overall response rate in the trial. There are many different algorithms such as random-play-the-winner (RPW), the utility-offset model and the maximum utility model.

5.1 Random-Play-the-Winner (RPW)



Response-Adaptive Randomization

Play-the-Winner:
When a response seen in yellow arm, yellow balls are added to the urn.

Randomization:
- Random select a color ball from the urn to determine the patient's treatment assignment.
- When a response seen in a treatment arm, balls of the corresponding color are added to the urn.
- Therefore more patients will be randomized into efficacious arms.

Fig. 2. Random-Play-the-Winner

The generalized RPW denoted by $RPW(n_1,\ n_2, ...,\ n_k;\ m_1,\ m_2, ...,\ m_k)$ can be described as follows.

(i) Place n$_i$ balls of the ith color (corresponding to the ith treatment) into a urn ($i = 1, 2, ..., k$), where k is number of treatment groups. There are initially $N = \sum n_i$ balls in the urn.

(ii) Randomly choose a ball from the urn. If it is the ith color, assign the next patient to the ith treatment group.

(iii) Add m$_k$ balls of the ith color to the urn for each response observed in the ith treatment. This creates more chances for choosing the ith treatment.

(iv) Repeat Steps (ii) and (iii).

When $n_i = n$ and $m_i = m$ for all i, we simply write $RPW(n, m)$ for $RPW(n_1,\ n_2, ...,\ n_k;\ m_1,\ m_2, ...,\ m_k)$.

5.2 Utility-Offset Model (UOM)

To have a high probability of achieving target patient distribution among the treatment groups, the probability of assigning a patient to a group should be proportional to the corresponding predicted or observed response rate minus the proportion of patients that have been assigned to the group.

5.3 Maximum Utility Model (MUM)

Maximum utility model for the adaptive-randomization always assigns the next patient to the group that has the highest response rate based on current estimation of either the observed or model-based predicted response rate.

6 Null Model versus Model Approach

It is interesting to compare model and null-model approaches. When sample size is larger than 20 per group, there is no obvious advantage by using the model-based method with respect to the precision and accuracy (Table 1). Therefore, null-model approach will be used in the subsequent simulations.

Dose Level	1	2	3	4	5
Target rate	0.02	0.07	0.37	0.73	0.52
Simulated rate	0.02	0.07	0.36	0.73	0.52
Predicted rate	0.02	0.07	0.40	0.65	0.41
Standard deviation	0.00	0.02	0.11	0.09	0.04
Number of subjects	1.02	2.48	12.6	20.5	13.4

Table 1. Comparisons of Simulations Results

7 Test Statistic

It is very interesting to know that the choice of test statistics for hypothesis tests is very flexible if the analysis is carried out through computer simulations. The only requirement is that the test statistic should be a monotonic function of both the treatment effect δ and sample size n, which can be, for example, $\sqrt{n}\delta$ the treatment difference or the effect size $\frac{\sqrt{n}\delta}{\sigma}$, where σ is the standard deviation of δ. Using computer simulation, it is easy to generate the distributions of the test statistic under the null hypothesis and alternative hypothesis or any other specified conditions for the monitoring purpose.

8 Bias in Rate Estimation and Alpha Adjustment

The commonly used estimators that are based on the assumption of independent samples are often biased in the case of adaptive design. The bias could be as much as 20

The α-adjustment is required when (i) there are multiple comparisons with more than two groups are involved, (ii) There are interim looks, i.e., early stopping for futility or efficacy, or (iii) There is a response-dependent sampling procedure such as response-adaptive randomization and unblinded sample size re-estimation. When samples or observations from the trial are not independent, the response data is no longer normally distributed. Therefore, the p-value from a normal distribution assumption should be adjusted or equivalently the alpha should be adjusted if the p-value is not adjusted. For the same reason, the other statistic estimates from normal assumption should also be adjusted.

9 Simulation Examples

To investigate the effect of the adaptations, we will compare the classic, group sequential and adaptive designs with regards to their operating characteristics using computer simulations. In what follows, each example represents a different trial design. All simulations are performed using ExpDesign Studio (www.CTriSoft.net) [CTriSoft, Intl. 2005]

Examples 1 to 3 will use the following scenario: Assume a phase II oncology trial with two treatment groups¿ The primary endpoint is tumor response (PR and CR) and the estimated response rates for the two groups are 0.2 and 0.3 respectively. We use simulation to calculate the sample size required, given that one-sided alpha = 0.05 and power = 80%.

Example 1: Conventional Design with Two Parallel Treatment Groups

A classic fixed sample size design with 600 subjects will have a power of 81.4% at one-sided $\alpha = 0.025$. The total number of responses per trial is 150 based on 10,000 simulations.

Example 2: Flexible design with Sample Size Re-estimation

Power of a trial is heavily dependent on the estimated effect size; therefore it is desirable to have a design that allows modification of sample size at some point during the trial. Let us re-design the trial in example 1 such that it allows a sample size-re-estimation and then study the robustness of the design.

In order to control the family-wise error rate (FWE) at 0.025, the alpha must be adjusted to 0.023 which can be obtained by computer simulation under the null hypothesis. The average sample size is 960 under the null hypothesis. Using the algorithm for sample size re-estimation (1), where $E_{0\max} = 0.1633$ and $a = 2$, the design has 92% power with an average sample size of 821.5.

Now assume the initial effect sizes are not 0.2 versus 0.3 for the two treatment groups. Instead, they are 0.2 and 0.28 respectively. We want to know what the power of the flexible design pertains. Keep everything the same (Also keep Eo_max 0.1633), but change the response rates to 0.2 and 0.28 for the two dose levels and run the simulation again. It turns out that the design has 79.4% power with an average sample size of 855.

Given the two response rates 0.2 and 0.28, the design with a fixed sample size of 880 has a power of 79.4%. We can see that there is a saving of 25 patients by using the flexible design. If the response rates are 0.2 and 0.3, for 92.1% power, the required sample size is 828 with the fixed sample size design, which means that the flexible design saves 6-7 subjects. A flexible design increases power when observed effect size is less than expected, while a traditional design with a fixed sample size either increases or decreases the power regardless of the observed effect size when the sample increases.

Example 3: Adaptive Design Permitting Early Stopping and Sample Size Re-estimation

It is some time desirable to have a design permitting both early stopping and sample size modification.

With an initial sample size of 700 subjects, a grouping size of 350, and a maximum sample size of 1000. The one-sided adjusted alpha is found to be 0.05. The simulation results are presented in the following.

The maximum sample size is 700. The trial will stop if 350 or more are randomized and one of the following criteria is met. (1) The efficacy (utility) stopping criterion: The maximum difference in response rate between any dose and the control is larger than 0.1 with the lower bound of the two-sided 95% naive confidence interval larger than or equal to 0.0. (2) The futility stopping criterion: The maximum difference in response rate between any dose and the control is smaller than 0.05 with the upper bound of the one-sided 95% naive confidence interval smaller than 0.1. The sample size will be re-estimated at the time when there are 350 subjects randomized.

When the null hypothesis is true ($p_1 = p_2 = 0.2$), the average total number of subjects for each trial is 398.8. The probability of early stopping for efficacy is 0.0096. The probability of early stopping for futility is 0.9638.

When the alternative hypothesis is true ($p_1 = 0.2,\ p_2 = 0.3$), the average total number of subjects for each trial is 543.5. The total number of responses per trial is 136. The probability of correctly predicting the most responsive dose level is 0.985 based on observed rates. The probability of early stopping for efficacy is 0.6225. The probability of early stopping for futility is 0.1546. The power for testing the treatment difference is 0.842.

Examples 4 to 6 are for the same scenario of the six arm study with response rates 0.5, 0.4, 0.5, 0.6, 0.7, and 0.55 for the 6 dose levels from dose 1 to 6, respectively.

Example 4: Conventional Design with Multiple Treatment Groups

With 800 subjects, 0.5 response rate under Ho, and grouping size of 100, we found the one-sided adjusted α to be 0.0055. The total number of responses per trial is 433. The probability of correctly predicting the most responsive dose level is 0.951 based on observed rates. The power for testing the maximum effect comparing any dose level to the control is 80%. The powers for comparing each of the 5 dose levels to the control are 0, 0.008, 0.2, 0.796, and 0.048, respectively.

Example 5: Response-Adaptive Design with Multiple Treatment Groups

To further investigate the effect of Random-Play-the-Winner randomization RPW(1,1), a design with 800 subjects, grouping size of 100, and a response rate of 0.2 under null hypothesis is simulated. The one-sided adjusted α is found to be 0.016. Using this adjusted alpha and response rates 0.5, 0.4, 0.5, 0.6, 0.7, and 0.55 for the dose levels 1 to 6, respectively, the simulation indicates that design trial has 86% power and 447 responders per trial on average. In comparison to 80% power and 433 responders for the design with simple randomization RPW(1,0), the adaptive randomization is superior in both power and number of responders. The simulation results also indicate there are biases in the estimated mean response rates in all dose levels except dose level 1, where a fixed randomization rate is used.

Dose level	1	2	3	4	5	6
Response rate	0.5	0.5	0.5	0.5	0.5	0.5
Observed rate	0.50	0.49	0.49	0.49	0.49	0.49

Table 2. Design with RPW(1,1) under H_o

Dose level	1	2	3	4	5	6
No. of subjects	200	74	100	133	176	116
Response rate	0.5	0.4	0.5	0.6	0.7	0.55
Observed rate	0.50	0.39	0.49	0.59	0.7	0.54

Table 3. Design with RPW(1,1) under H_a

The average total number of subjects for each trial is 800. The total number of responses per trial is 446.8. The probability of correctly predicting the most responsive dose level is 0.957 based on observed rates. The power for testing the maximum effect comparing any dose level to the control (dose level 1) is 0.861 at a one-sided significant level (alpha) of 0.016. The powers for comparing each of the 5 dose levels to the control are 0, 0.008, 0.201, 0.853, and 0.051, respectively.

Example 6: Adaptive Design with Dropping Losers

Implementing the mechanism of dropping loser can also improve the efficiency of a design. With 800 subjects, grouping size of 100, a response rate of 0.2 under the null hypothesis, and fixed randomization rate in dose level 1 at 0.25, an inferior group (loser) will be dropped if the maximum difference in response between the most effective group and the least effective group (loser) is larger than 0 with the lower bound of the one-sided 95% naive confidence interval larger than or equal to 0. Using the simulation, the adjusted alpha is found to be 0.079. From the simulation results below, more biases can be observed with this design. The design has 90% power with 467 responders. The probability of correctly predicting the most responsive dose level is 0.965 based on observed rates. The powers for comparing each of the 5 dose levels to the control (Dose level 1) are 0.001, 0.007, 0.205, 0.889, and 0.045, respectively. The design is superior to both RPW(1,0) and RPW(1,1).

Dose level	1	2	3	4	5	6
Response rate	0.5	0.5	0.5	0.5	0.5	0.5
Observed rate	0.50	0.46	0.46	0.46	0.46	0.46

Table 4. Bias in Rate with dropping losers under H_o

10 Summary

From classic design to group sequential design to adaptive design, each step forward has an increased complexity and at the same time improves the efficiency of clinical trials. Adaptive design can increase the number of responses

Dose level	1	2	3	4	5	6
No. of subjects	200	26	68	172	240	95
Response rate	0.5	0.4	0.5	0.6	0.7	0.55
Observed rate	0.50	0.37	0.46	0.57	0.69	0.51

Table 5. Bias in Rate with dropping losers under H_a

in a trial and provide more benefits to the patient in comparison to the classic design. With sample size re-estimation, an adaptive design can preserve the power even when the initial estimations of treatment effect and its variability are inaccurate. In the case of a multiple-arm trial, dropping inferior arm or response-adaptive randomization can improve the efficiency of a design dramatically. Finding analytic solutions for adaptive designs is theoretically challenging. However, computer simulation makes it easier to achieve an optimal adaptive design. It allows a wide range of test statistics as long as they are monotonic functions of treatment effects. Adjusted alphas and p-values due to response-adaptive randomization and other adaptations with multiple comparisons can be easily determined using computer simulations. Unbias in point estimation with adaptive design has not completely revolved yet using computer simulations. However, the bias can be ignored in practice by using a proper grouping size (cluster) such that there are only a limited number of adaptations (< 8).

References

[Bauer and Kohne, 1994]Bauer, P. and Kohne, K. (1994). Evaluation of experiments with adaptive interim analyses. Biometrics 50, 1029-1041. Correction in Biometrics 52, 380.

[Babb and Rogatko, 2004]Babb, J.S. and Rogatko, A. (2004). Bayesian methods for cancer phase I clinical trials, Advances in Clinical Trial Biostatistics, Nancy L. Geller (ed.), Marcel Dekker, Inc, 2004.

[CTriSoft, Intl. 2005]CTriSoft, Intl. (2005). ExpDesign Studio Manual, Lexington, MA, USA.

[Cui et al., 1999]Cui, L., Hung, H.M.J. and Wang, S.J. (1999). Modification of sample size in group sequential clinical trials. Biometrics 55, 853-857.

[Lan and Demets, 1988]Lan, K.K.G. and Demets D.L. (1988), Discrete sequential boundaries for clinical trials. Biometriku (1988), 70, 3, pp, 659-663.

[Lin et al., 1999]Lin, D.Y., Tao, Q. and Ying, Z. (1999), A general theory on stochastic curtailment for censored survival data. JASA, (1999) Vol. 94, No. 446.

[O'Quigley et al., 1990]O'Quigley, J., Pepe, M., and Fisher, L. (1990). Continual reassessment method: A practical design for phase I clinical trial in cancer, Biometrics 46:33-48.

[Proschan and Hunsberger 1995]Proschan, M.A. and Hunsberger, S.A. (1995). Design extension of studies based on conditional power. Biometrics 51, 1315-1324.

[Rosenberger and Lachin, 2002]Rosenberger, W.F. and Lachin J.M. (2002). Randomization in Clinical Trials, John Wiley & Sons, Inc., New York.